

PAT-NO: JP409107534A

DOCUMENT-IDENTIFIER: JP 09107534 A

TITLE: VIDEO CONFERENCE EQUIPMENT AND VIDEO CONFERENCE
SYSTEM

PUBN-DATE: April 22, 1997

INVENTOR-INFORMATION:

NAME

SATO, HAJIME

TANAKA, ATSUSHI

YANAGISAWA, RYOZO

KOBAYASHI, KATSUYUKI

YOSHIMURA, YUICHIRO

TOKIOKA, MASAKI

ASSIGNEE-INFORMATION:

NAME

COUNTRY

CANON INC

N/A

APPL-NO: JP07263174

APPL-DATE: October 11, 1995

INT-CL (IPC): H04N007/15, H04N005/232

ABSTRACT:

PROBLEM TO BE SOLVED: To provide a video conference equipment and a video conference system by which video conference is held with excellent naturalness as if a participant were actually making conversation with opposite conference party face to face.

SOLUTION: A position detection section 6 detects a position of an opposite image display area 10 on a display section 3 and sends the position information to a camera control section 2. An image recognition section 7 recognizes an image of an opposite party displayed on the opposite image display area 10 and sends position information representing a face or eye position of the person to

the camera control section 2. A user position detection section detects a position of an utterance party from the position of a microphone picking up a voice of the utterance party and the position information is sent to the camera control section 2. The camera control section 2 is configured to move the position of a camera 1 in three axis directions XYZ and the camera 1 is moved so that the position of the camera 1 comes to a position reaching the line of sight of the user depending on the position information obtained from the position detection section 6, the image recognition section 7 and a user position detection section 9 or the like.

COPYRIGHT: (C)1997,JPO

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-107534

(43) 公開日 平成9年(1997)4月22日

(51) Int.Cl. [°]	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 N 7/15			H 0 4 N 7/15	
5/232			5/232	B

審査請求 未請求 請求項の数 4 O L (全 9 頁)

(21) 出願番号 特願平7-263174

(22) 出願日 平成7年(1995)10月11日

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 佐藤 肇

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72) 発明者 田中 淳

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72) 発明者 柳澤 亮三

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(74) 代理人 弁理士 國分 孝悦

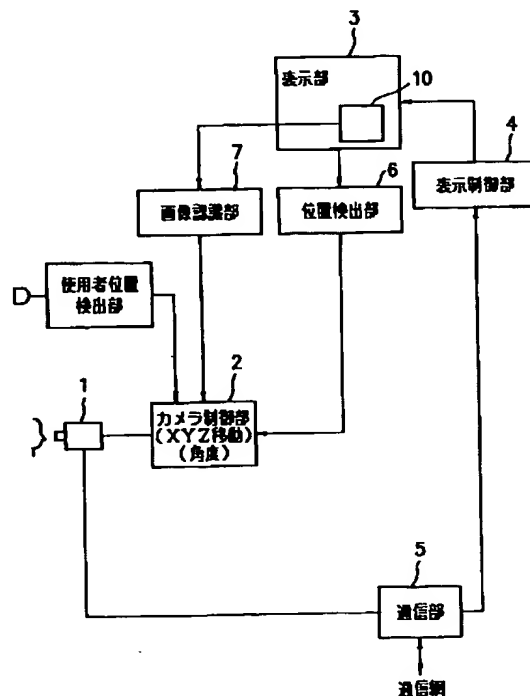
最終頁に続く

(54) 【発明の名称】 テレビ会議装置及びテレビ会議システム

(57) 【要約】

【課題】 会議の相手と実際に対面して会話を行ってような自然性に優れたテレビ会議を行うことができるテレビ会議装置及びテレビ会議システムを提供する。

【解決手段】 位置検出部6は、表示部3上における相手画像表示領域10の位置を検出し、その位置情報をカメラ制御部2に送出する。画像認識部7は、相手画像表示領域10に表示される相手の人物の画像を認識し、その人物の顔あるいは目の位置を示す位置情報をカメラ制御部2に送出する。使用者位置検出部9は、発言者の音声が入力されたマイクロフォン8の位置から発言者の位置を検出し、その位置情報をカメラ制御部2に送出する。カメラ制御部2は、カメラ1の位置をXYZの3軸方向に移動させることができるように構成され、位置検出部6、画像認識部7、使用者位置検出部9等から得られる位置情報に応じてカメラ1の位置が使用者の目線の先位置にくるようにカメラ1を移動させる。



【特許請求の範囲】

【請求項1】 テレビ会議に用いるテレビ会議装置において、

画像を表示する表示手段と、

使用者を撮影する撮影手段と、

テレビ会議の相手側の画像を上記表示手段の相手画像表示領域に表示させる表示制御手段と、

上記表示手段における上記相手画像表示領域の位置を検出する位置検出手段と、

上記位置検出手段により検出された上記相手画像表示領域の位置に応じて上記撮影手段の位置を変化させる撮影制御手段とを有することを特徴とするテレビ会議装置。

【請求項2】 テレビ会議に用いるテレビ会議装置において、

画像を表示する表示手段と、

使用者を撮影する撮影手段と、

テレビ会議の相手側の画像を上記表示手段に表示させる表示制御手段と、

上記相手側の画像の画像認識を行う画像認識手段と、

上記画像認識手段によって認識された上記相手側の画像の所定部分の表示位置に応じて上記撮影手段の位置を変化させる撮影制御手段とを有することを特徴とするテレビ会議装置。

【請求項3】 テレビ会議に用いるテレビ会議装置において、

画像を表示する表示手段と、

使用者を撮影する撮影手段と、

発言を行った使用者の位置を検出する使用者位置検出手段と、

上記使用者位置検出手段によって検出された発言を行った使用者の位置に応じて上記撮影手段の位置又は角度を変化させる撮影制御手段とを有することを特徴とするテレビ会議装置。

【請求項4】 多地点間通信を行ってテレビ会議を行うテレビ会議システムにおいて、

画像を表示する表示手段と、

自分側の画像を撮影する撮影手段と、

テレビ会議の相手側の画像を上記表示手段の相手画像表示領域に表示させる表示制御手段と、

上記表示手段における上記相手画像表示領域の位置を検出する位置検出手段と、

上記位置検出手段により検出された上記相手画像表示領域の位置に応じて上記撮影手段の位置を変化させる撮影制御手段とを有することを特徴とするテレビ会議システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、テレビ会議に用いるテレビ会議装置及び多地点間通信を行ってテレビ会議を行うテレビ会議システムに関するものである。

【0002】

【従来の技術】 従来のテレビ会議システムは、図7に示すように、パーソナルコンピュータ（パソコン）やワークステーション703に映像・音声コーデックが内蔵されたシステム制御装置704及びカメラ701、モニター702、マイク705、ビデオボード706、LANボード707等の周辺機器を付加して構成されたテレビ会議装置をLAN等の通信網を介して接続することにより構成されている。そしてこのテレビ会議システムは、映像・音声の通信だけでなく、マルチウィンドウ、マルチタスク、データ蓄積、電子メール等、パソコンやワークステーション703の機能と複合させたマルチメディア通信システムが構成できることが特徴である。

【0003】ところで、このテレビ会議システムでは、ヒューマンファクタが技術的に占める部分も多く、利用者の心理的要因を考慮したヒューマンインタフェースの設計が行われている。そしてテレビ会議システムは通常、複数の参加者による意見交換に用いられるので、臨場感、刺激感といったものが重要な要素になってくる。

【0004】この臨場感、刺激感を感じるためには、会議を行っている相手とのアイコンタクトが必要になるため、従来のテレビ会議装置では、相手先の画面と使用者の視線のずれを考慮して、使用者を撮影するカメラ701はモニター702の上部に配置されていることが多い。これは、上記視線ずれは、上下（垂直）方向の許容度の方が水平方向の許容度より大きいためである。

【0005】

【発明が解決しようとする課題】 しかしながら、上記のように構成した場合でも、特に大型のディスプレイ（モニター）を用いたテレビ会議装置の場合には、視線ずれの許容度が小さいため、使用者が意識してカメラ701に視線を投じたとき以外は、モニター702に出る相手先の顔画像は視線がずれていることが多い。従って、このような構成のテレビ会議装置を用いてテレビ会議を行っても、人が普通に会話を行っているような自然さに欠け、目の前に相手がいるような臨場感、刺激感を得ることができない。

【0006】 本発明は、上記の問題を解決するためになされたものであり、実際に対面して会話を行っているような自然性に優れたテレビ会議を行うことができるテレビ会議装置及びテレビ会議システムを提供することを目的とする。

【0007】

【課題を解決するための手段】 請求項1の発明は、テレビ会議に用いるテレビ会議装置において、画像を表示する表示手段と、使用者を撮影する撮影手段と、テレビ会議の相手側の画像を上記表示手段の相手画像表示領域に表示させる表示制御手段と、上記表示手段における上記相手画像表示領域の位置を検出する位置検出手段と、上記位置検出手段により検出された上記相手画像表示領域

の位置に応じて上記撮影手段の位置を変化させる撮影制御手段とを有することを特徴とするものである。

【0008】請求項2の発明は、テレビ会議に用いるテレビ会議装置において、画像を表示する表示手段と、使用者を撮影する撮影手段と、テレビ会議の相手側の画像を上記表示手段に表示させる表示制御手段と、上記相手側の画像の画像認識を行う画像認識手段と、上記画像認識手段によって認識された上記相手側の画像の所定部分の表示位置に応じて上記撮影手段の位置を変化させる撮影制御手段とを有することを特徴とするものである。

【0009】請求項3の発明は、テレビ会議に用いるテレビ会議装置において、画像を表示する表示手段と、使用者を撮影する撮影手段と、発言を行った使用者の位置を検出する使用者位置検出手段と、上記使用者位置検出手段によって検出された発言を行った使用者の位置に応じて上記撮影手段の位置又は角度を変化させる撮影制御手段とを有することを特徴とするものである。

【0010】請求項4の発明は、多地点間通信を行ってテレビ会議を行うテレビ会議システムにおいて、画像を表示する表示手段と、自分側の画像を撮影する撮影手段と、テレビ会議の相手側の画像を上記表示手段の相手画像表示領域に表示させる表示制御手段と、上記表示手段における上記相手画像表示領域の位置を検出する位置検出手段と、上記位置検出手段により検出された上記相手画像表示領域の位置に応じて上記撮影手段の位置を変化させる撮影制御手段とを有することを特徴とするものである。

【0011】

【作用】請求項1のテレビ会議装置及び請求項4のテレビ会議システムによれば、撮影制御手段により例えば表示手段における相手画像表示領域の位置と同じ位置に撮影手段を移動させるようにすることにより、使用者が相手と会話するために相手画像表示領域に目を投じると、自然に撮影手段を見ることになるので、相手と視線を合わせて対話を行っているかのような自然性に優れたテレビ会議を行うことができる。

【0012】請求項2のテレビ会議装置によれば、画像認識手段によって認識された相手の画像の例えば顔又は目の位置に撮影制御手段により撮影手段を移動させるようにすることにより、特に大型ディスプレイを用いた場合でも、使用者が相手と会話するために相手の画像を見ると自然に撮影手段を見ることになるので、相手と視線を合わせることなく相手と視線を合わせて対話を行っているかのような自然性に優れたテレビ会議を行うことができる。

【0013】請求項3のテレビ会議装置によれば、使用者位置検出手段により発言者の位置を検出し、その位置に応じて撮影制御手段により撮影手段の位置又は角度を変化させることができるので、自機器側の使用者が複数いる場合でも、発言者の方に撮影手段を向けたり、発言

者の正面に撮影手段を移動させることができ、従って相手と対面して会話を行っているかのような自然性に優れたテレビ会議を行うことができる。

【0014】

【発明の実施の形態】以下、本発明の実施例を図面に基づき説明する。図8は、本発明の要素的特徴を示すブロック図である。同図において、通信部5は通信網を介して相手側のテレビ会議装置との間で画像、音声、データ等の送受信を行う。そして通信部5で受信された画像は表示制御部4により表示部3に表示される。

【0015】一方、撮影手段を構成するカメラ1により撮影された使用者（自分側）の画像は、通信部5及び通信網を介して相手側のテレビ会議装置に送信される。

【0016】表示部3には、種々の情報が表示されると共に、会議相手の画像が表示部3の一部に設けられた相手画像表示領域10に表示されるように構成されている。そしてこの相手画像表示領域10は、表示部3上で任意の位置に移動させることができ、また、その表示領域の大きさを変化させることができるようになされている。

【0017】位置検出部6は、表示部3上における相手画像表示領域10の位置を検出し、その位置情報をカメラ制御部2に送出する。

【0018】画像認識部7は、相手画像表示領域10に表示される相手の人物の画像を認識し、その人物の顔あるいは目の位置を示す位置情報をカメラ制御部2に送出する。

【0019】使用者位置検出部9は、発言者の音声が入力されたマイクロフォン8の位置から発言者の位置を検出し、その位置情報をカメラ制御部2に送出する。

【0020】カメラ制御部2は、カメラ1の位置をXYZの3軸方向に移動させることができるように構成され、位置検出部6、画像認識部7、使用者位置検出部9等から得られる位置情報に応じてカメラ1の位置が使用者の視線の先の位置にくるようにカメラ1を移動させる。

【0021】なお、カメラ制御部2は、位置検出部6、画像認識部7、使用者位置検出部9等から得られる位置情報に応じてカメラ1の角度も自由に变化させることができるように構成されている。

【0022】なお、上述の表示制御部4、位置検出部6、画像認識部7、使用者位置検出部9は、後述するシステム制御装置109、409により実現される。

【0023】〔第1実施例〕図1は、本発明の第1の実施例のテレビ会議装置の構成を示す図である。同図において、109はこのテレビ会議装置全体を制御するシステム制御装置である。図示しないマイクで検出する音声とカメラ104で撮影される映像は、システム制御装置109の内部の映像・音声コーデックで処理され、ビデオボード等を介してパソコンやワークステーションか

らなるコンピュータ108に入力される。そしてこのコンピュータ108、システム制御装置109が、LAN或いはISDN等の通信手段を介して他のテレビ会議装置と接続されることによりテレビ会議システムが構築され、多地点での映像、音声、情報等の通信が可能となる。

【0024】101は透明な部材で構成されるデジタイザであり、ペン型の指示具で座標入力を行うことができるように構成され、デジタイザ制御装置107で制御される。

【0025】102は透過型液晶表示装置であり、この液晶表示装置102には端末装置のモニタとして各種のデータが表示される。

【0026】103はハーフミラーであり、使用者に対するカメラ104の映り込みを避けるのと同時に、LAN等の通信手段で接続されたシステム制御装置109に設けられている映像コーデックで処理される画像が投写装置106により投写される。この投写される画像は主にテレビ会議の相手のポートレート画像であり、使用者は液晶表示装置102及びデジタイザ101を通してこの画像を映像として認識することができる。

【0027】コンピュータ108の制御により液晶表示装置102に表示される各種情報の映像と、投写装置106で投写される映像は、重なって使用者に認識されないように、システム制御装置109によりその表示位置等が制御される。

【0028】カメラ104は、カメラ制御装置105によりその撮影位置及び撮影条件等の制御がなされ、投写装置106がハーフミラー103に対して投写する映像の領域、すなわちポートレート画像表示ウィンドウ領域（相手画像表示領域）の表示位置に応じてその移動制御がなされる。カメラ104の移動は、XYZステージ等と通称される3次元の位置決め移動装置で行われ、前述のポートレート画像表示ウィンドウが位置する座標値にその位置制御がなされる。

【0029】図2は、カメラ104と、投写装置106が投写するポートレート画像表示ウィンドウ領域が連動して動作する様子を示したものである。なお、ここではポートレート画像表示ウィンドウ領域の大きさは固定とする。

【0030】使用者がデジタイザ101の操作により、液晶表示装置102及びハーフミラー103で構成される画面201中のポートレート画像表示ウィンドウ領域を202の位置から203の位置に移動させると、カメラ104はカメラ制御装置105の制御により、このポートレート画像表示ウィンドウ領域の移動に連動して205から206の位置に移動される。

【0031】従って、使用者がポートレート画像表示ウィンドウ領域に目を投じたときには、カメラ104を見ていることになる。

【0032】図3は、拡大あるいは縮小できるポートレート画像表示ウィンドウ領域に対してカメラ104の位置制御がなされる様子を示したものである。

【0033】上記ポートレート画像表示ウィンドウ領域は、使用者のペン或いはマウス等の操作により任意に拡大、縮小できるように構成されているが、ポートレート画像表示ウィンドウ領域の表示位置の移動がなされると共にその領域の大きさの拡大、縮小がなされる場合には、カメラ制御装置105によりカメラ104の位置の以下のような移動制御がなされる。

【0034】まず、ポートレート画像表示ウィンドウ領域がカメラ104程度の大きさであればカメラ104はポートレート画像表示ウィンドウ領域の中心位置に移動される。

【0035】また、ポートレート画像表示ウィンドウ領域がカメラ104よりも大きな場合には、システム制御装置109に設けられた画像認識部によりポートレート画像表示ウィンドウ領域内の人物の画像の画像認識を行い、その顔あるいは目の位置にカメラ104を移動させる。すなわち、図3に示すように、使用者が画面301の中のポートレート画像表示ウィンドウ領域を302の位置から303の位置に移動させると共に、ポートレート画像表示ウィンドウ領域の拡大を行うと、カメラ104はカメラ制御装置105により305の位置から306の位置に移動される。この306の位置は、上述のように画像認識により特定された相手の人物の顔あるいは目の位置である。従って、大型のディスプレイ（画面）を使用している場合のように、使用者の視線の動きが大きくなるような場合にも相手のポートレート画像を見たときにはカメラ104を見ているようになる。このように構成することにより、画像認識は投写装置106によって投写される映像のみに対して行えば良いので、液晶表示装置102に表示される画像の処理部には負荷がかからない。そして、使用者は相手先の画像を見ると自然にカメラ104を見ることになるので、実際に対面して会話をするような自然性に優れたテレビ会議システムを提供することができる。

【0036】〔第2実施例〕上述の第1実施例では、相手先の画像を表示する表示部（ポートレート画像表示ウィンドウ領域）とデータ等を表示する表示部（液晶表示装置102）とは別々に構成されていた。本実施例では、相手先の画像とデータ等を同一の表示処理により表示させる場合の例につき説明する。なお、図1の実施例と同一部分には同一の符号を付してその説明を省略する。

【0037】図4において、402は液晶制御装置406によりその表示が制御される透過型液晶表示装置で、システム制御装置409で一括処理される各種のデータおよび相手先の画像を表示する。403は使用者へのカメラ104の映り込みを防止するための部材である。

【0038】カメラ104の位置は、液晶表示装置402に表示された画像の内容に応じてカメラ制御装置105で制御される。すなわち、システム制御装置409では、液晶表示装置402に表示された画像から相手先の画像のウィンドウ（相手画像表示領域）を認識し、さらにそのウィンドウ領域内の画像から、相手の人物の顔あるいは目を画像認識する。そして認識された顔あるいは目の位置情報はカメラ制御装置105に入力され、カメラ制御装置105はその位置情報に基づきカメラ104の位置を相手先の画像のウィンドウに表示された会議相手の顔あるいは目の位置にカメラ104の位置を移動させる。

【0039】このように、この実施例では、表示画面（液晶表示装置402）全体の画像認識を段階を踏んで実行し、その実行結果に基づいてカメラ104の移動制御を行うことで、使用者は表示装置402の画面上で相手先の画像を見ると自然にカメラ104を見ることとなるので、実際に対面して会話をするような自然性に優れたテレビ会議システムを提供することができる。

【0040】〔第3実施例〕次に、図5、図6に基づき、複数の地点の相手と複数の使用者がテレビ会議を行う場合の例につき説明する。

【0041】この場合には、図6に示すように、複数の地点の相手の画像が表示される各々のポートレート画像表示ウィンドウ領域に対応してカメラ制御装置105及びカメラ104を複数配置すると共に、自機器側の複数の使用者の人数分のマイク1、2、3・・・を各使用者に対応して配置する。なお、図6において、「表示側」とあるのは、図5の画面501側を表し、「使用者側」とあるのは、図5の510で示される使用者が座っている場所を表している。

【0042】この構成において、ある使用者が発言すると、その発言した使用者に対応させられているマイクにより音声が出検され、その検出情報がシステム制御装置109（409）に送られる。システム制御装置109（409）は、予め登録されている各マイクの位置情報に基づきその発言者の位置を特定し、その位置情報を、画像認識により認識したポートレート画像表示ウィンドウ領域の人物の顔又は目の位置情報と共にカメラ制御装置105に送出する。

【0043】カメラ制御装置105は、それらの位置情報を用いて、各カメラ104をポートレート画像表示ウィンドウ領域の人物の顔又は目の位置に移動させると共に、カメラ104の角度を変化させ、カメラ104を発言者の方向に向けるように制御する。なお、この際、カメラ104の位置は固定で、発言者の位置情報のみを用いてカメラ104の角度のみを変化させるようにしても良い。

【0044】図5は、図6の構成を具体的に説明するための図である。図5に示すように、画面501には、

ラフ等のデータを表示するウィンドウ502、503が表示されると共に、テレビ会議の相手であるA地点、B地点、C地点の各々の会議参加者のポートレート画像がポートレート画像表示ウィンドウ領域504、505、506に表示されている。そしてこのウィンドウ領域504、505、506の裏側にはそれぞれカメラ104が配置されている。

【0045】一方、自機器側の使用者はA～Cの3名であり、この3名は図5の510に示すような配置で座っており、このA～Cの使用者の前にはそれぞれマイクA～Cが配置されているものとする。

【0046】今、使用者Cが発言したとすると、マイクCが使用者Cが発生する音声を検出する。この検出に基づき、システム制御装置109（409）は、予め登録されている各マイクの位置情報に基づきその発言者Cの位置を特定し、その位置情報を、画像認識により認識したポートレート画像表示ウィンドウ領域の人物の顔又は目の位置情報と共にカメラ制御装置105に送出する。カメラ制御装置105は、それらの位置情報を用いて、各カメラ104をポートレート画像表示ウィンドウ領域の人物の顔又は目の位置に移動させると共に、カメラ104の角度を変化させ、各カメラ104を矢印507、508、509で示すように発言者Cの方向に向けるように制御する。

【0047】このように、自機器側に複数の使用者がいる場合に、発言を行った使用者にカメラ104が向けられるので、実際に対面して会話をするような自然性に優れたテレビ会議システムを提供することができる。

【0048】なお、図5の例では相手先はA、B、Cの3地点であったが、これは何地点であっても良く、また、カメラ104を各ポートレート画像表示ウィンドウ領域に対応してそれぞれ設けるのではなく、1台のカメラ104のみを用いてこのカメラ104を発言を行った使用者の正面の位置に移動させたり、発言を行った使用者の方向に向けたりするように構成しても良い。

【0049】上述の実施例では、パソコン或いはワークステーションにカメラ104等を組み合わせて本発明のテレビ会議装置を構成する例につき説明したが、パソコン或いはワークステーションの代わりにCPU等を用いた専用機で本発明のテレビ会議装置を構成しても良い。

【0050】また、上述の第1及び第2の実施例において、カメラ104の移動後、カメラ制御部105によりカメラ104の角度も自動的に変え、使用者の視線がカメラ104のレンズに正対するように構成しても良い。このようにすることにより、使用者と相手との視線のずれが全くなり、より一層自然な対話を行うことが可能となる。

【0051】

【発明の効果】以上説明したように、本発明のテレビ会議装置及びテレビ会議システムによれば、会議の相手と

10

20

30

40

50

実際に対面して会話を行っているような自然性に優れたテレビ会議を行うことができる。

【図面の簡単な説明】

【図1】本発明の第1の実施例に係るテレビ会議装置の構成を示すブロック図である。

【図2】本発明の第1の実施例におけるカメラの移動制御例を示す図である。

【図3】本発明の第1の実施例におけるカメラの移動制御例を示す図である。

【図4】本発明の第2の実施例に係るテレビ会議装置の構成を示すブロック図である。

【図5】本発明の第3の実施例におけるカメラの制御例を説明するための図である。

【図6】本発明の第3の実施例に係るテレビ会議装置の構成を示すブロック図である。

【図7】従来のテレビ会議システムの構成を示した図である。

【図8】本発明の要素的特徴を示すブロック図である。

【符号の説明】

1 カメラ（撮影手段）

2 カメラ制御部（撮影制御手段）

3 表示部

4 表示制御部

5 通信部

6 位置検出部

7 画像認識部

8 マイク

9 使用者位置検出部

10 相手画像表示領域

101 デジタイザ

102 透過型液晶表示装置

103 ハーフミラー

104 カメラ

105 カメラ制御装置

106 投写装置

107 デジタイザ制御装置

108 コンピュータ

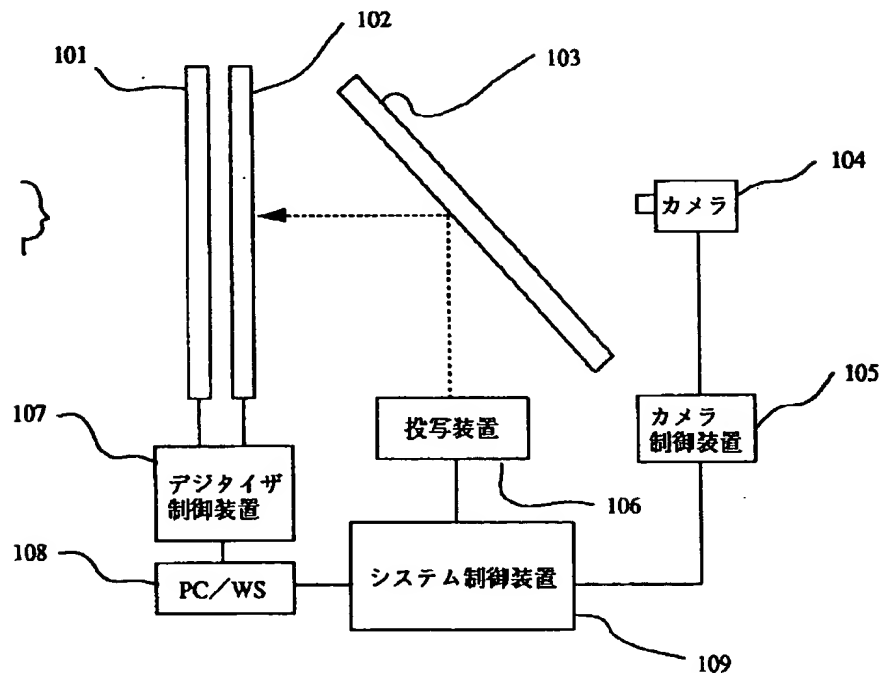
109 システム制御装置

402 透過型液晶表示装置

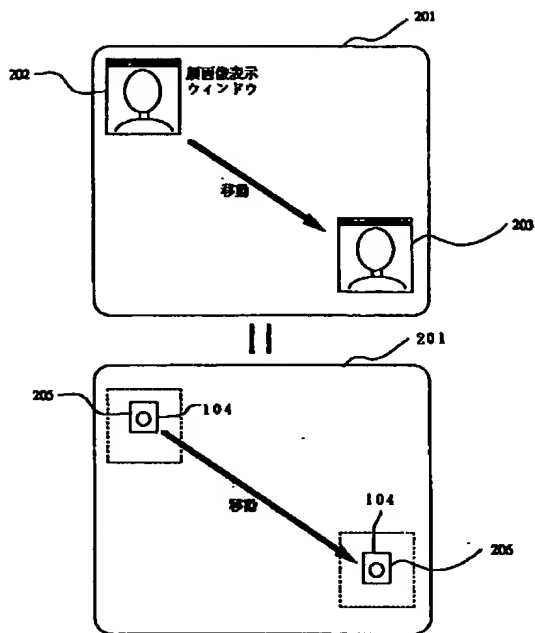
406 液晶制御装置

20 409 システム制御装置

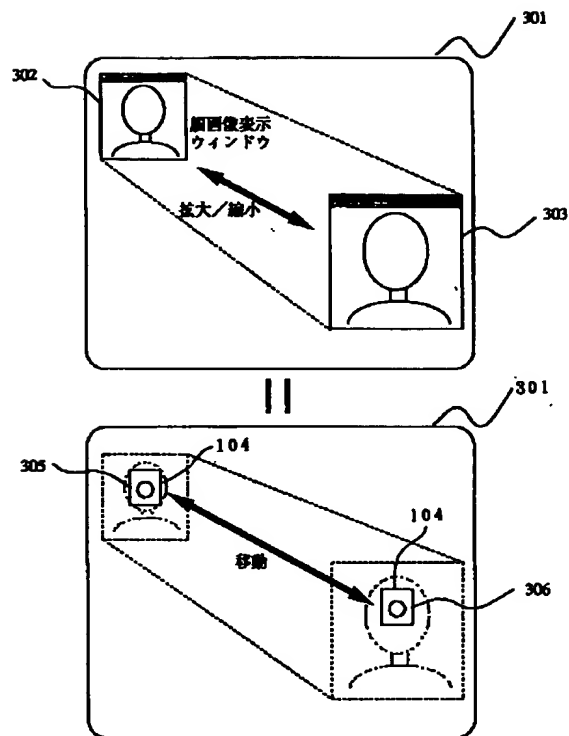
【図1】



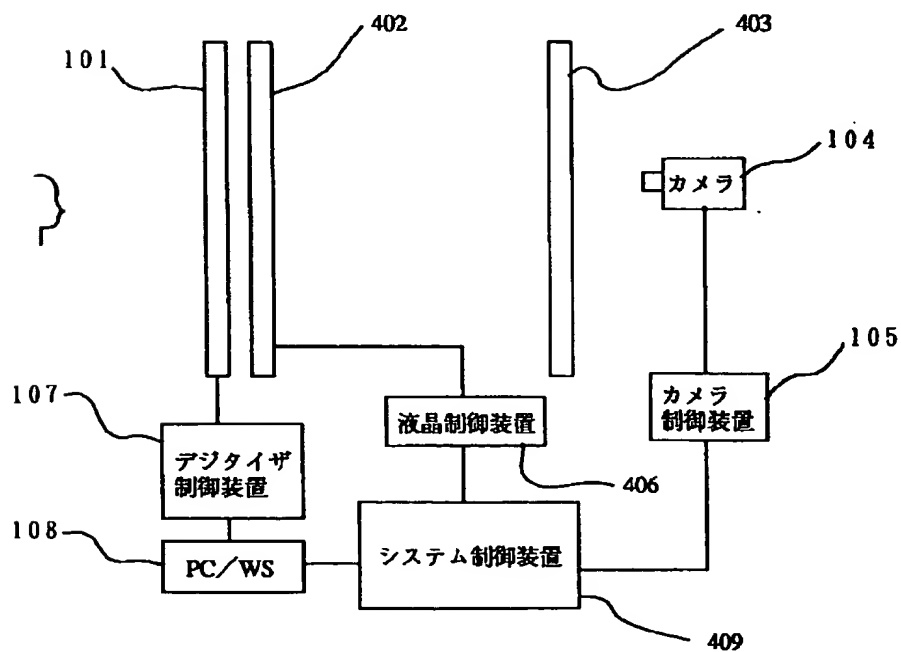
【図2】



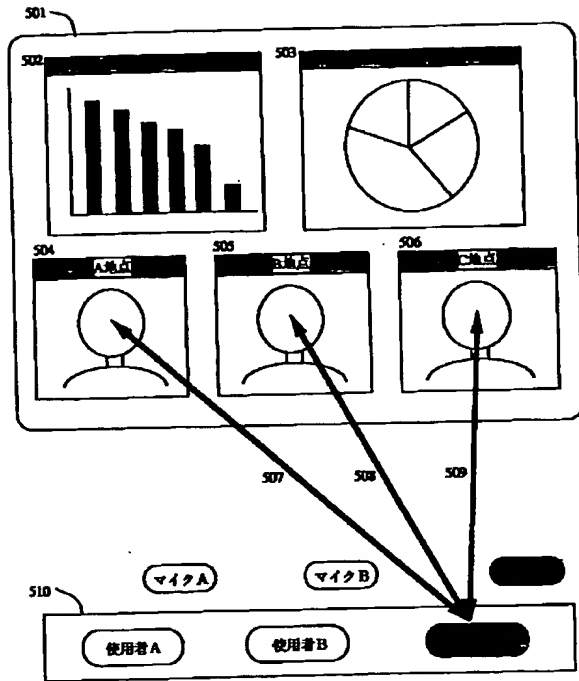
【図3】



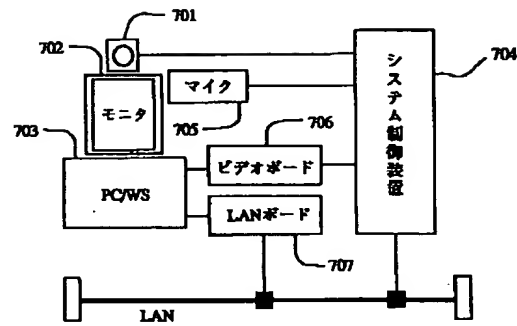
【図4】



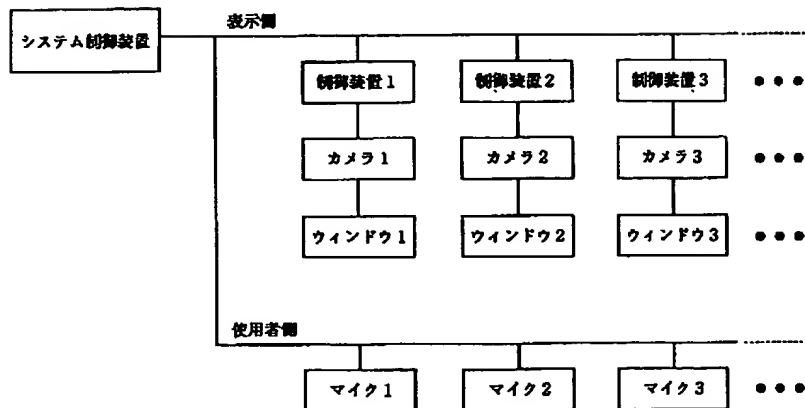
【図5】



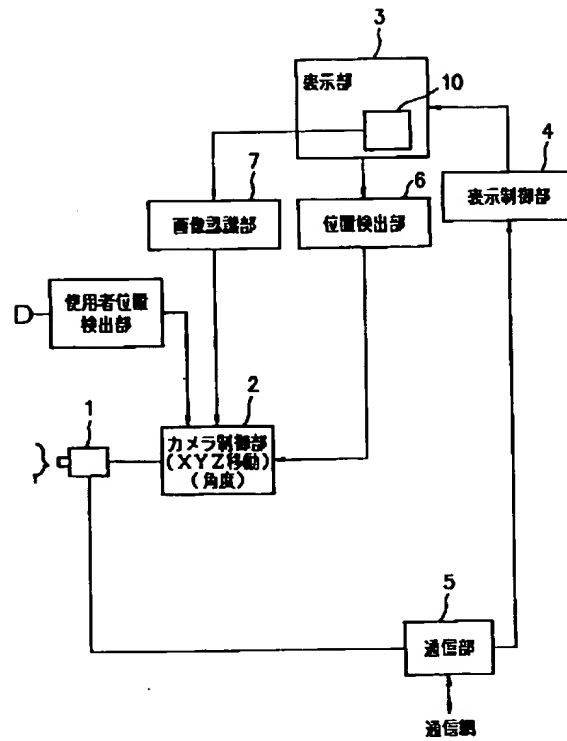
【図7】



【図6】



【図8】



フロントページの続き

(72)発明者 小林 克行
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(72)発明者 吉村 雄一郎
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内
(72)発明者 時岡 正樹
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the video conference system which performs the television conference equipment used for a television conference, and the communication link between many points, and performs a television conference.

[0002]

[Description of the Prior Art] The conventional video conference system is constituted by connecting to a personal computer (personal computer) or a workstation 703 the television conference equipment constituted by adding the peripheral device of the system control station 704 with which the image and the voice codec were built in and a camera 701, a monitor 702, a microphone 705, a video board 706, and LAN board 707 grade through communication networks, such as LAN, as shown in drawing 7. And it is the description that this video conference system can constitute the multimedia communication system compounded with the function of personal computers, such as not only the communication link of an image and voice but a multi window, multitasking, data accumulation, an electron mail, etc., or a workstation 703.

[0003] By the way, in this video conference system, there are also many parts which a human factor occupies technically, and the design of the human interface in consideration of a user's psychological factor is performed. And since a video conference system is usually used for the opinion exchange by two or more participants, things, such as a presence and a feeling of a stimulus, become an important element.

[0004] Since eye contact with the partner who is holding a conference in order to sense this presence and a feeling of stimulus is needed, with conventional television conference equipment, the camera 701 which photos a user is arranged in consideration of the gap of a phase hand's screen and a user's look in the upper part of a monitor 702 in many cases. This is because the above-mentioned look gap is larger than tolerance with the more nearly horizontal tolerance of the vertical (perpendicular) direction.

[0005]

[Problem(s) to be Solved by the Invention] However, in the case of the television conference equipment using an especially large-sized display (monitor), even when constituted as mentioned above, since the tolerance of a look gap is small, except when a user is conscious and an eye line is invested in a camera 701, as for the face image of the phase hand who appears in a monitor 702, the eye line is shifted in many cases. Therefore, even if it performs a television conference using the television conference equipment of such a configuration, nature with which people are talking ordinarily is missing, and the presence and the feeling of a stimulus a partner is before an eye cannot be obtained.

[0006] This invention is made in order to solve the above-mentioned problem, and it aims at offering the television conference equipment and the video conference system which can perform the television conference excellent in natural manner which is talking by actually meeting.

[0007]

[Means for Solving the Problem] A display means to display an image in the television conference equipment which uses invention of claim 1 for a television conference, A photography means to photo a user, and a display-control means to display the image of the other party of a television conference on the partner image display field of the above mentioned display means, It is characterized by having a location detection means to detect the location of the above mentioned partner image display field in the above-mentioned display means, and the photography control means to which the location of the above-mentioned photography means is changed according to the location of the above-mentioned partner image display field detected by the above-mentioned location detection means.

[0008] A display means to display an image in the television conference equipment which uses invention of claim 2

a television conference, A photography means to photo a user, and a display-control means to display the image of the other party of a television conference on the above-mentioned display means, It is characterized by having an image recognition means to perform the image recognition of the image of the above-mentioned other party, and the photography control means to which the location of the above-mentioned photography means is changed according to the display position of the predetermined part of the image of the above-mentioned other party recognized by the above-mentioned image recognition means.

[0009] Invention of a claim 3 carries out having a display means display an image, the photography means photo a user, the user location detection means detect the location of the user who performed an utterance, and the photograph control means, to which the location or the include angle of the above-mentioned photography means changes according to the location of the user who performed the utterance detected by the above-mentioned user location detection means in the television conference equipment which uses for a television conference as the description.

[0010] In the video conference system which invention of claim 4 performs the communication link between many points, and performs a television conference A display means to display an image, a photography means to photo the image by the side of oneself, and a display-control means to display the image of the other party of a television conference on the partner image display field of the above-mentioned display means, It is characterized by having a location detection means to detect the location of the above-mentioned partner image display field in the above-mentioned display means, and the photography control means to which the location of the above-mentioned photography means is changed according to the location of the above-mentioned partner image display field detected by the above-mentioned location detection means.

[0011]

[Function] According to the television conference equipment of claim 1, and the video conference system of claim 4 By the photography control means, for example, by making it move a photography means to the same location as the location of the partner image display field in a display means Since a photography means will be automatically seen when an eye is invested in a partner image display field, in order that a user may talk with a partner, the television conference excellent in that natural gender [like] which is having a dialog by doubling an eye line with a partner can be performed.

[0012] By making it move a photography means to the location of a partner's image, for example, a face, recognized the image recognition means, or an eye by the photography control means according to the television conference equipment of claim 2 Since a photography means will be automatically seen when especially a large-sized display is used, and a partner's image is seen, in order that a user may talk with a partner, the television conference excellent in that natural gender [like] which is having a dialog by doubling an eye line with a partner can be performed without producing a look gap with a partner.

[0013] Since according to the television conference equipment of claim 3 a user location detection means can detect speaker's location and the location or include angle of a photography means can be changed by the photography control means according to the location Even when there are two or more users by the side of a self-device, the television conference excellent in that natural gender [like] which can turn a photography means to speakers, can be made to move a photography means to a speaker's transverse plane, therefore is talking by meeting a partner can be performed

[0014]

[Embodiment of the Invention] Hereafter, the example of this invention is explained based on a drawing. Drawing 8 the block diagram showing the element-description of this invention. In this drawing, the communications department transmits an image, voice, data, etc. and receives between the television conference equipment of the other party through a communication network. And the image received in the communications department 5 is displayed on a display 3 by the display and control section 4.

[0015] A user's (oneself side) image photoed on the other hand with the camera 1 which constitutes a photography means is transmitted to the television conference equipment of the other party through the communications department 5 and a communication network.

[0016] While various information is displayed, it is constituted by the display 3 so that it may be displayed on the partner image display field 10 to which a meeting partner's image was prepared in a part of display 3. And it can be made to move to the location of arbitration on a display 3, and this partner image display field 10 is made as [change the magnitude of that viewing area].

[0017] The location detecting element 6 detects the location of the partner image display field 10 on a display 3, and sends out the positional information to the camera control section 2.

[0018] The image recognition section 7 recognizes partner's person's image displayed on the partner image display field 10, and sends out the positional information which shows the location of the person's face, or an eye to the camera

control section 2.

[0019] The user location detecting element 9 detects a speaker's location from the location of the microphone 8 into which a speaker's voice was inputted, and sends out the positional information to the camera control section 2.

[0020] The camera control section 2 is constituted so that the location of a camera 1 can be moved to 3 shaft orientations of XYZ, and it moves a camera 1 so that the location of a camera 1 may come to the location of the point user's eye line according to the positional information acquired from the location detecting element 6, the image recognition section 7, and user location detecting-element 9 grade.

[0021] In addition, the camera control section 2 is constituted so that the include angle of a camera 1 can also be free changed according to the positional information acquired from the location detecting element 6, the image recognition section 7, and user location detecting-element 9 grade.

[0022] In addition, the above-mentioned display and control section 4, the location detecting element 6, the image recognition section 7, and the user location detecting element 9 are realized by the system control stations 109 and 4 mentioned later.

[0023] The [1st example] Drawing 1 is drawing showing the configuration of the television conference equipment of the 1st example of this invention. In this drawing, 109 is a system control station which controls this whole television conference equipment. The voice detected with the microphone which is not illustrated and the image photoed with a camera 104 are processed by the image and voice codec inside a system control station 109, and is inputted into the computer 108 which consists of a personal computer or a workstation through a video board etc. And by connecting this computer 108 and a system control station 109 with other television conference equipments through means of communications, such as LAN or ISDN, a video conference system is built and the communication link of the image many points, voice, information, etc. is attained.

[0024] 101 is a digitizer which consists of transparent members, it is constituted so that the directions implement of a pen mold can perform a coordinate input, and it is controlled by the digitizer control unit 107.

[0025] 102 is a transparency mold liquid crystal display, and various kinds of data are displayed on this liquid crystal display 102 as a monitor of a terminal unit.

[0026] 103 is a half mirror and the image processed by the image codec prepared in the system control station 109 connected by means of communications, such as avoiding reflected [to a user / a camera 104], simultaneously LAN projected by the projection arrangement 106. This image projected is mainly a portrait image of the partner of a television conference, and a user can recognize the image of a through lever for a liquid crystal display 102 and a digitizer 101 as an image.

[0027] The display position etc. is controlled by the system control station 109 so that the image of the various information displayed on a liquid crystal display 102 by control of a computer 108 and the image projected with a projection arrangement 106 lap and is not recognized by the user.

[0028] The migration control is made according to the display position of the field of an image where control of the camera station, photography conditions, etc. is made by the camera control unit 105, and a projection arrangement 10 projects a camera 104 to a half mirror 103, i.e., a portrait image display window area, (partner image display field). Migration of a camera 104 is performed by the positioning migration equipment of the three dimension by which a common name is carried out to a XYZ stage etc., and the position control is made by the coordinate value with which the above-mentioned portrait image display window is located.

[0029] Drawing 2 shows signs that the portrait image display window area which a projection arrangement 106 projects is interlocked with a camera 104, and it operates. In addition, magnitude of a portrait image display window area is considered as immobilization here.

[0030] If a user moves the portrait image display window area in Screen 201 which consists of a liquid crystal display 102 and a half mirror 103 to the location of 203 from the location of 202 by actuation of a digitizer 101, a camera 10 will be interlocked with migration of this portrait image display window area, and will be moved to the location of 2 to 206 by control of the camera control unit 105.

[0031] Therefore, the camera 104 will be seen when a user invests an eye in a portrait image display window area.

[0032] Drawing 3 shows signs that the position control of a camera 104 is made to the portrait image display window area which can be expanded or reduced.

[0033] The above-mentioned portrait image display window area is constituted so that it can expand and reduce to arbitration by actuation of a user's pen or a mouse, but while migration of the display position of a portrait image display window area is made, when expansion of the area size and contraction are made, the following migration control of the location of a camera 104 is made by the camera control unit 105.

[0034] First, if a portrait image display window area is the magnitude which is about 104 camera, a camera 104 will

moved to the center position of a portrait image display window area.

[0035] Moreover, when a portrait image display window area is bigger than a camera 104, the image recognition section prepared in the system control station 109 performs the image recognition of the image of the person in a portrait image display window area, and a camera 104 is moved to the location of the face or an eye. That is, if a portrait image display window area is expanded while a user moves the portrait image display window area in Screen 301 to the location of 303 from the location of 302 as shown in drawing 3, a camera 104 will be moved to the location of 306 by the camera control unit 105 from the location of 305. This location of 306 is a location of partner's person's face specified according to image recognition as mentioned above, or an eye. Therefore, when a motion of a user's eye line becomes large and a partner's portrait image is seen like [in the case of using the large-sized display (screen)], it is coming to see the camera 104. Thus, since what is necessary is just to carry out by constituting only to the image on which image recognition is projected by the projection arrangement 106, a load is not applied to the processing section of the image displayed on a liquid crystal display 102. And since a user will look at a camera 104 automatically when he looks at a phase hand's image, he can offer the video conference system excellent in natural gender which talks by actually meeting.

[0036] The [2nd example] The display (portrait image display window area) which displays a phase hand's image, and the display (liquid crystal display 102) which displays data etc. consisted of the 1st above-mentioned example separately. This example explains a phase hand's image, data, etc. per example in the case of making it display by the same display processing. In addition, the same sign is given to the same part as the example of drawing 1, and the explanation is omitted.

[0037] In drawing 4, 402 is the transparency mold liquid crystal display with which the display is controlled by the liquid crystal control unit 406, and displays various kinds of data by which batch processing is carried out with a system control station 409, and a phase hand's image. 403 is a member for preventing reflected [the camera 104 to a user].

[0038] The location of a camera 104 is controlled by the camera control unit 105 according to the contents of the image displayed on the liquid crystal display 402. That is, in a system control station 409, the window (partner image display field) of a phase hand's image is recognized from the image displayed on the liquid crystal display 402, and the image recognition of a partner's face or eye of a person is further carried out from the image in the window area. And the positional information of the recognized face or an eye is inputted into the camera control unit 105, and the camera control unit 105 moves the location of a camera 104 to the location of a meeting partner's face displayed on the window of a phase hand's image in the location of a camera 104 based on the positional information, or an eye.

[0039] Thus, in this example, a phase is stepped on, the image recognition of the whole display screen (liquid crystal display 402) is performed, and by performing migration control of a camera 104 based on that activation result, since a user will look at a camera 104 automatically when he looks at a phase hand's image on the screen of a display 402, he can offer the video conference system excellent in natural gender which talks by actually meeting.

[0040] The [3rd example] Next, based on drawing 5 and drawing 6, it explains per example in case the partner of two or more points and two or more users perform a television conference.

[0041] In this case -- as shown in drawing 6, while arranging two or more camera control units 105 and cameras 104 corresponding to each portrait image display window area where the image of the partner of two or more points is displayed -- the microphones 1, 2, and 3 for the number of two or more users by the side of a self-device .. is arranged corresponding to each user. In addition, in drawing 6, that it is a "display side" expresses the Screen 501 side of drawing 5, and that it is a "user side" expresses the location where the user shown by 510 of drawing 5 is sitting down.

[0042] In this configuration, if a certain user speaks, voice will be detected by the microphone made to correspond to that user that spoke, and that detection information will be sent to a system control station 109 (409). A system control station 109 (409) pinpoints the speaker's location based on the positional information of each microphone registered beforehand, and sends it out to the camera control unit 105 with the face of the person of a portrait image display window area who has recognized the positional information according to image recognition, or the positional information of an eye.

[0043] The camera control unit 105 changes the include angle of a camera 104, and is controlled to turn a camera 104 in the direction of a speaker while moving each camera 104 to the location of the face of the person of a portrait image display window area, or an eye using those positional information. In addition, the location of a camera 104 is immobilization and you may make it change only the include angle of a camera 104 only using a speaker's positional information in this case.

[0044] Drawing 5 is drawing for explaining the configuration of drawing 6 concretely. As shown in drawing 5, while the windows 502 and 503 which display data, such as a graph, are displayed on Screen 501, the portrait image of each

meeting participant in A point which is the partner of a television conference, B point, and C point is displayed on the portrait image display window areas 504, 505, and 506. And the camera 104 is arranged on the background of these window areas 504, 505, and 506, respectively.

[0045] On the other hand, the user by the side of a self-device shall be the trinomial of A-C, this trinomial shall sit down by arrangement as shown in 510 of drawing 5, and microphone A-C shall be arranged in front of the user of the A-C, respectively.

[0046] Now, supposing User C speaks, Microphone C will detect [User C] the voice to generate. Based on this detection, a system control station 109 (409) pinpoints that speaker's C location based on the positional information of each microphone registered beforehand, and sends it out to the camera control unit 105 with the face of the person of portrait image display window area who has recognized that positional information according to image recognition, or the positional information of an eye. The camera control unit 105 changes the include angle of a camera 104, and is controlled to turn each camera 104 in the direction of Speaker C, as arrow heads 507, 508, and 509 show while moving each camera 104 to the location of the face of the person of a portrait image display window area, or an eye using the positional information.

[0047] Thus, since a camera 104 is turned to the user who spoke when two or more users are in a self-device side, the video conference system excellent in natural gender which talks by actually meeting can be offered.

[0048] In addition, although phase hands were three points of A, B, and C in the example of drawing 5 This may be any point and does not form a camera 104 corresponding to each portrait image display window area, respectively. Y may constitute so that it may be made to move to the location of the transverse plane of the user who spoke or this camera 104 may be turned in the direction of the user who spoke only using one camera 104.

[0049] Although explained per [which constitutes the television conference equipment of this invention combining camera 104 grade to a personal computer or a workstation] example, the special-purpose machine which used CPU instead of the personal computer or the workstation may constitute the television conference equipment of this invention from an above-mentioned example.

[0050] Moreover, in the 1st and 2nd above-mentioned examples, the include angle of a camera 104 may also be automatically changed by the camera control section 105 after migration of a camera 104, and you may constitute so that a user's eye line may carry out a right pair to the lens of a camera 104. By doing in this way, a gap of the look of user and a partner is completely lost and becomes possible [performing a much more natural dialogue].

[0051]

[Effect of the Invention] As explained above, according to the television conference equipment and the video conference system of this invention, the television conference excellent in natural gender which is talking by actually meeting the partner of a meeting can be performed.

[Translation done.]